

# 웹 크롤링과 이미지 색상 추출 인공지능을 이용한 향수 추천 서비스 구현

김유진<sup>1</sup>, 이예림<sup>2</sup>, 정성윤<sup>3</sup>, 조유진<sup>4</sup>, 김영종<sup>5</sup>

<sup>1</sup> 숭실대학교 소프트웨어학부 학부생

<sup>2</sup> 숭실대학교 소프트웨어학부 학부생

<sup>3</sup> 숭실대학교 소프트웨어학부 학부생

<sup>4</sup> 숭실대학교 소프트웨어학부 학부생

<sup>5</sup> 숭실대학교 소프트웨어학부 교수

zzqaws38@naver.com, redheadannielim@gmail.com, simpler0407@naver.com,  
soyjo0508@naver.com, opensys@soongsil.ac.kr

## Implementation of perfume recommendation service using web crawling and image color extraction artificial intelligence

Yu-jin Kim<sup>1</sup>, Ye-lim Lee<sup>2</sup>, Sung-Yoon Jung<sup>3</sup>, Yu-jin Jo<sup>4</sup>

<sup>1</sup>Dept. of Software, Soong-Sil University

<sup>2</sup>Dept. of Software, Soong-Sil University

<sup>3</sup>Dept. of Software, Soong-Sil University

<sup>4</sup>Dept. of Software, Soong-Sil University

### 요 약

이 논문에서는 웹 크롤링과 인공지능의 색상 추출 기능을 사용하여 사용자에게 맞는 향수를 추천해주는 서비스를 구현한다. 웹 사이트 제작에 용이한 Java 와 웹 크롤링과 인공지능 구현에 용이한 Python 을 기반으로 구현하였다.

### 1. 서론

코로나19가 감소세에 접어들며 하나 둘 마스크를 벗기 시작하자, 후각을 다루는 시장들이 다시금 활기를 띠고 있다. 그 예로 '향수'를 꼽을 수 있다 [1]사회 심리학자들에 따르면 후각은 사람의 행동과 다양한 심리적인 작용에 영향을 주며 대뇌로 직접 전달되기 때문에 기억에 오래 남는다. 사람들은 향수를 사용하여 자신의 이미지를 들어내며 지위, 나이, 성별, 성격, 스타일에 따라 각자 자신에게 맞는 향수를 선택하게 된다.

대체로 향수 사용자들은 자신의 이미지와 맞는 향을 구매하고 싶어한다. 그러나 모든 종류의 향을 직접 맡아보는 것은 불가능하고, 대부분의 사람들은 전문적으로 잘 알지 못하기 때문에 자신의 이미지와 향수를 고르는 데에 어려움을 느낀다. 따라서 웹

크롤링과 인공지능을 이용해 이미지에 맞는 향수를 추천해주고자 한다.

### 2. 본론

#### 2.1 크롤링을 이용한 키워드 추출

사용자의 이미지 데이터를 통해 향수에 이미지를 부여하기 위한 방법으로 크롤링(Crawling)을 택하였다. 필요한 것은 향에 대한 '객관적인 텍스트' 정보였으며, 이를 기대하기에 가장 적합한 방법이라고 여겼다.

웹 페이지 소스로는 뷰티 랭킹 어플 및 웹 '글로우픽(Glow Pick)' 을 활용하였다. 글로우픽은 향수, 화장품, 생활용품 등 다방면의 아이템을 다루며 소비자의 '리뷰' 에 초점이 맞춰져있다. 구글스토어 4.5의 별점과, 40만 명의 사용자를 보유하고 있어 데이터의 신뢰도를 보장할 수 있다고

판단하였다. 향수 데이터는 2023년 3월 XX일을 기준으로 상위 20개의 향수 데이터를 사용하였다.

향수 하나당 50 개의 리뷰를 추출하였다. 텍스트 데이터이므로 시각화의 필요성이 요구되었고, 직관적 판단에 용이한 워드 클라우드(word cloud)를 사용하였다.

2.2 인공지능을 이용한 색상추출

앞서 데이터 크롤링 기법을 사용하여 얻은 결과를 토대로 각 향수별로 어울리는 이미지를 찾을 수 있게 되었다. 이 이미지들과, 사용자가 원하는 분위기를 가진 이미지의 색상을 추출하고 비교하여 가장 적절한 향수를 추천해줄 수 있는 알고리즘을 개발하였다.

알고리즘의 과정은 다음과 같다. 먼저, 향수별로 어울리는 이미지를 대표하는 색상을 추출한다. 사용자 이미지도 같은 과정을 거친다. 그 다음, RGB 값들을 list 에 등록한다. 마지막으로, 사용자 RGB 값과 가장 가까운 향수 이미지의 RGB 값을 찾고 사용자에게 해당 향수를 추천해준다.

3. 구현

3.1 크롤링을 이용한 키워드 추출

[2]selenium 을 이용하여 해당 웹 페이지 html 의 class 로 원하는 리뷰 데이터를 수집한다.

수집한 데이터를 txt 파일로 변형하여 명사와 형용사를 추출한다. 불용어(stopword)로는 향수 각각의 특성을 나타낸다고 보기 어려운 무의미한 단어와 단순한 긍정을 내포하는 단어를 선택한다.

.wordcloud 와 matplotlib 를 사용하여, 리뷰 데이터 중 출현 빈도가 높은 단어 80 개를 기준으로 시각화한다.

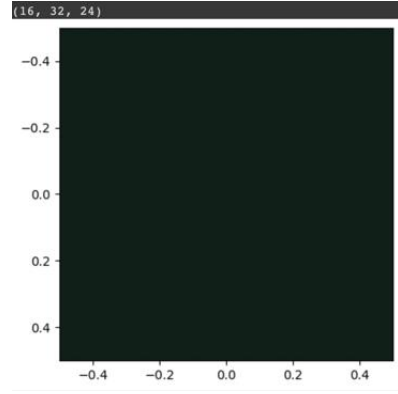


<그림 1> 크롤링을 이용한 워드클라우드 결과

3.2 인공지능을 이용한 색상추출

이미지에서 색상을 추출하기 위해 Python 라이브러리 중 ColorThief 을 사용하였다. 또한, 추출한 색상을 시각적으로 보여주기 위해 Matplotlib 이 필요하다. 사용자 이미지에서 추출한 색상과 향수 리스트에

등록되어 있는 이미지 색상을 비교하여 가장 적절한 향수를 사용자에게 추천해준다.



<그림 2> 추천 향수에 해당하는 이미지의 RGB 값

4. 향후 연구 방향

본 논문은 인공지능과 데이터 크롤링 기법을 사용하여 소비자와 어울리는 이미지 키워드를 통해 향수를 추천하는 서비스를 제안한다.

일일히 정보검색을 해야 하는 기존 서비스들과 달리 유형별 이미지에 어울릴 만한 향수를 한번에 쉽게 확인할 수 있다. 또한 이미지가 비슷한 이용자들끼리 분류해 서로의 스타일과 그 사람들이 사용하는 향수를 확인할 수 있다.

향후 인공지능의 발전과 더불어 이미지에 어울리는 향수 추천 시스템도 산업적, 학문적으로 더 발전이 이루어질 것으로 기대된다.

참고문헌

[1] 민병찬.인체에 미치는 향의 심리 / 생리적 효과. 한국정밀공학회지,18(2),46-53.(2001)  
 [2] 김근교. "뉴스 웹페이지 크롤링과 텍스트 정제 및 중요단어 추출 함수구현." 국내석사학위논문 아주대학교, 2022.